DOCUMENT RESUME

ED 415 964 PS 026 131

AUTHOR Valente, Ernest, Jr.; Dodge, Kenneth A.
TITLE Evaluation of Prevention Programs for Children.
SPONS AGENCY National Inst. of Mental Health (DHHS), Bethesda, MD.; Department of Education, Washington, DC.; Substance Abuse and Mental Health Services Administration (DHHS/PHS), Rockville, MD. Center for Substance Abuse Prevention.
PUB DATE 1997-00-00
NOTE 37p.; In: Establishing Preventive Services. Healthy Children 2010. Issues in Children's and Families' Lives, Vol. 9; see PS 026 124.
CONTRACT R18MH48083; R18MH50951; R18MH50952; R18MH50953; K05MH01027
PUB TYPE Reports - Descriptive (141)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Adolescents; *Child Health; Child Welfare; Children; *Early Intervention; Health Programs; Health Promotion; Health Services; *Mental Health; *Mental Health Programs; *Prevention; Program Effectiveness; *Program Evaluation
IDENTIFIERS *Healthy Children 2010; Healthy People 2000; Service Quality

ABSTRACT

This article focuses on the formal evaluation of large-scale preventive interventions promoting positive mental health in children and adolescents, using examples of conduct problems. The state of the art in program evaluation is discussed based on quantitative evaluations of mental illness prevention programs. The article reviews current thinking in three related areas of evaluation of health improvement programs for children. First, a case is made for the necessity of evaluations of mental health improvement programs, including a discussion of recent trends in evaluation research in general. Second, more focus specifically on prevention research is encouraged. The dual role of evaluation in the prevention context as both testing efficacy and information theory is presented. Finally, a discussion concerning specifics about design and statistical issues raised by this perspective is presented. The article concludes by addressing how the presented approach to evaluation of prevention programs addresses stages in the prevention research cycle: (1) theory development, which allows for further theoretical understanding of illness processes; (2) implementation clarity, which allows for enhancement of effective, large-scale prevention programs; (3) identification of effective programs to ensure maximum impact; (4) more effective disseminated programs to enhance the likelihood of success; and (5) grounds for improvement to assist in the refinement and success of subsequent preventive interventions. Contains approximately 100 references. (SD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 415 964

PS 026131



PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
James Lee Henson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

# Evaluation of Prevention Programs for Children

ERNEST VALENTE, JR.
KENNETH A. DODGE

A mid-decade review of the national policy plan that was begun 5 years ago (called *Healthy People 2000*; U.S. Department of Health and Human Services, Public Health Service [DHHS], 1991b) expresses some optimism that lives of children in the United States are improving. The report also indicates that despite this good progress, significant challenges remain, especially in the prevention of behavioral maladjustment, mental disorder, and violence (DHHS, 1995; Fleming, 1996). One important challenge is the long-term prevention of these problems. In 1994, the Institute of Medicine (IOM) issued a guide for the next decade of research in preventive intervention (IOM, 1994). This guide suggested that research in prevention in any domain follows a five-stage sequence, with a feedback loop from the last stage to the first stage indicating that the process is iterative. The five stages are (a) identification of the problem, its prevalence, and lifecourse; (b) identification of risk and protective factors for the problem so that

AUTHORS' NOTE: We acknowledge support from the following sources: NIMH Grants R18MH48083, R18MH50951, R18MH50952, and R18MH50953 awarded to the Conduct Problems Prevention Research Group (Karen L. Bierman, John D. Coie, Kenneth A. Dodge, Mark T. Greenberg, John E. Lochman, and Robert J. McMahon); Department of Education Grant B5184430002, the Center for Substance Abuse Prevention; and Research Scientist Award K05MH01027.

BEST COPY AVAILABLE

its development can be understood; (c) design and analysis of innovative pilot interventions to prevent the problem; (d) the design, implementation, and analysis of large-scale trials of a preventive intervention program; and (e) large-scale implementation and dissemination of the program to society. This chapter is directed toward the fourth stage in this sequence, that is, the formal evaluation of large-scale preventive interventions promoting positive mental health in children and adolescents.

Although the concepts in this chapter apply to a broad array of substantive problems (Stage 1 of the IOM model), we will cite examples in the domain of conduct problems, clearly one of the most pressing issues facing society. Young children with conduct problems disrupt school learning environments. Untreated, early conduct problems can evolve into more serious, sometimes criminal, behaviors. Some children with conduct problems evolve into habitual, violent criminals. Homicide is already the major cause of death for young African American males, and all projections suggest that this trend will spread to young members of all demographic groups over the next decade. The treatment of violent behavior in adults is complicated by its intractability. Once established firmly in childhood and adolescence, it is resistant to change. It appears that long-term prevention attempts begun in early childhood hold the most potential for reducing violent crime (Yoshikawa, 1994).

It is a difficult task to design violence prevention programs that can be refined, established, and accepted in the relatively short period of time between now and 2010. Successful programs will incorporate cutting-edge theoretical advances into their design (at Stage 2 of the IOM model). Several recent papers have summarized the theoretical advancements in the field of conduct disorder (e.g., Coie & Dodge, in press; Hinshaw, 1994). Current theory suggests that adolescent conduct disorder is predictable from the time of early childhood, and that multiple factors play roles in its development, including biological predisposition, the sociocultural context in which a child lives, parenting quality, peer relationships, social cognitive skills, academic performance, and adult role models. These theoretical advancements have led to the design of interventions in diverse areas (Stage 3 of the IOM model), including parent training, academic tutoring, social cognitive skills training, classroom curricula, mentoring, and peer relationships enhancement





(Conduct Problems Prevention Research Group [CPPRG], 1992). Although each of these intervention targets has proven to lead to short-term positive changes in child functioning, the long-term preventive efficacy of any single intervention has not yet been proven. The most parsimonious explanation of this set of findings is that conduct problems development is overdetermined and can be perpetuated by influences in any one of the domains cited above. The time is ripe for comprehensive prevention efforts that incorporate interventions from all of the relevant developmental domains. Such a large-scale prevention trial is what the IOM calls the fourth stage of prevention research. Several prototypes are now undergoing field trials, including the Metropolitan Area Child Study (Guerra, Huesmann, Tolan, Van Acker, & Eron, 1995), the FAST Track Study (CPPRG, 1992), and the Baltimore preventive trial (Kellam, Rebok, Ialongo, & Mayer, 1994).

Program quality must be demonstrated if effective programs are to be recognized, accepted, and used. It is our intention to argue for the utility of program evaluation, in the context of field trials of candidate programs, in providing the information needed to achieve these goals. Although conducting a high-quality evaluation is a difficult task, the information gathered in evaluation research is critically important in the creation and dissemination of innovative programs (Weiss, 1995).

Our contribution in this chapter is to outline what we consider to be the state of the art in program evaluation and to discuss recent advances. Much can be written about evaluation, and we must decide on some organizing principles to present a coherent discussion. We have selected organizing principles that reflect both the purpose of this volume and our own beliefs about evaluation. Our discussion will therefore be limited to theory-based experimental or quasi-experimental quantitative evaluations of mental illness prevention programs for children and adolescents.

We have chosen to focus on theory-based evaluations because of their potential to provide both high-quality program effectiveness data and substantive information about the illness process that is the target of the treatment program. Theory-based evaluation designs are guided by the expectations about how and why the prevention program should work. These expectations are formed through causal beliefs, informed by contemporary basic theory, about the underlying illness processes. This close relation between

causal theory and evaluation design improves the validity of the evaluation findings (Chen & Rossi, 1987) and generates evaluation designs that can provide important information about the underlying causal processes (Cook, Anson, & Walchli, 1993).

We have a strong preference for randomized experimental or nonrandomized quasi-experimental quantitative evaluation research designs (Boruch & Wothke, 1985; Campbell & Stanley, 1963; Cook & Campbell, 1979). We do recognize the view that qualitative designs are useful in evaluation research (Cronbach, 1982; Guba & Lincoln, 1981; Lincoln & Guba, 1985; Pearsol, 1987), but this chapter reflects our bias toward quantitative studies. The advantages of randomized experiments are obvious in their potential for more definitive conclusions regarding the causal effectiveness of the intervention itself without the potential confounds and alternate explanations that are inherent in nonexperimental demonstrations. We believe that such demonstrations have tremendous value in publicizing and disseminating programs, which constitutes the last stage of the IOM model of prevention research strategy. A well-conducted experiment affords the high level of confidence in outcomes that is necessary to make a convincing case for wide acceptance and use of an innovative program.

We focus on the evaluation of prevention programs (rather than other intervention efforts) for several important reasons. The DHHS has emphasized prevention as a national health improvement strategy for adults (DHHS, 1991b) and children (*Healthy Children 2000*; DHHS, 1991a), making prevention programs the focus of much current research. The evaluation of theoretically derived prevention efforts is an important step in the refinement of subsequent prevention efforts (IOM, 1994; Price & Smith, 1985), and theoretically guided evaluations of prevention programs can help advance substantive knowledge about the underlying causal mechanisms that result in the illness that is the target of the prevention effort (Kellam & Rebok, 1992). Thus, the proliferation of prevention research and the role of evaluation in advancing knowledge of illness processes point to the need for evaluation in the prevention research process.

We are psychologists with experience in evaluation of mental illness prevention programs, so we will focus on mental health programs rather than physical health programs. Our examples will





be drawn from the mental illness prevention literature. We believe, however, that the concepts presented here will generalize to the evaluation of any kind of prevention effort.

It is our belief that theory-based evaluations of prevention efforts provide a unique opportunity both to improve subsequent intervention efforts and to advance substantive knowledge about illness etiology. Issues in theory-based evaluation of prevention programs also subsume issues common to evaluation in general, so this perspective provides a good framework to present recent advances in the broader field of evaluation science. It is therefore our goal to make a case for the benefit of theory-based evaluations of mental health prevention programs and to discuss some important technical and methodological aspects of this type of evaluation process.

The present chapter reviews current thinking in three related areas of evaluation of health improvement programs for children. First, we will make a brief case for the necessity of evaluations of mental health improvement programs, and we will discuss recent trends in evaluation research in general. Next, we will focus more specifically on prevention research, arguing for the dual role of evaluation in the prevention context as both testing efficacy and informing theory. Finally, we will discuss specifics about design and statistical issues raised by this perspective.

## The Emerging Field of Prevention Science

Evaluation has long been considered an important, but neglected, part of the process of designing effective preventive interventions (Bickman, 1983; Lorion, 1983; Lorion & Lounsbury, 1981). Evaluation in this context has been historically viewed primarily as a way to provide information about the efficacy of programs—an important way to determine program utility and to argue for the potential economic benefit of preventive efforts (Lorion, 1983). One example is the early evaluations of Project Head Start (House Select Committee on Children, Youth, and Families, 1988), which assessed the effectiveness of this early intervention in helping children from poverty backgrounds achieve success in early elementary school.

Recent advances in evaluation science in general, and evaluation of prevention programs in particular, have broadened the scope of

information that can be provided by evaluation research. Theory-based evaluation (Bickman, 1987b; Chen, 1991) can provide theoretically informative data, making evaluation an even more useful and pervasive part of the program design and dissemination process. In their critical review, Cook et al. (1993) concluded that high-quality theory-based evaluation research provides both efficacy information and theoretical information about the target of the program.

Concomitant with these advances in evaluation science, the role of evaluation has been more completely integrated into the overall prevention research process. In their guide to evaluating prevention programs, Price and Smith (1985) presented a model for the conduct of prevention research that is a useful framework for understanding the various activities required in the development of a prevention program. Their model shows how developing a useful prevention program requires the integration of several activities: exploration, theory development, evaluation, dissemination, and revision. In this view, evaluation is not only a way to demonstrate program efficacy but a central and important feature of program development.

Price and Smith (1985) described the conduct of prevention research as a process that encompasses four major domains of activity: (a) problem analysis, where the problem to be prevented is defined and described; (b) design innovation, where the knowledge of the underlying illness process is studied, and a suitable prevention is designed; (c) field trials, where the proposed prevention programs are evaluated; and (d) dissemination of successful prevention programs (or successful intervention components) to users. The more recent IOM model further subclassifies the first stage as involving separable stages of problem identification and theory development.

Price and Smith explain that each of the domains is within itself cyclical, so that each stage may be repeated until enough information is gathered to warrant progressing to the next stage. Information gathered at each stage is used at the next stage of the process. For example, information gathered at the problem analysis stage should provide information about the target population and risk factors that will guide the design of the prevention program.

Another important feature of the process is that the product of a successfully completed field trial cycle should be a treatment innovation or project prototype described in sufficient detail to be carried out by other agencies or groups. This ties the research process to application, defining realistic parameters for the provisional prevention program.





Findings from the evaluation of a proposed prevention program can feed back to the beginning of the process, informing substantive understanding of the illness process and the design of subsequent prevention programs. The importance of this feedback loop was further emphasized in a recent IOM report (1994) concerning the prevention of mental disorders.

The need for high-quality evaluation is also driven by the fact that many diverse intervention programs have been proposed to prevent identical mental health problems. Some of these programs are complementary (such as parent training and child social cognitive skills training), but other sets of programs may seem (especially to the public) contradictory. Consider, for example, juvenile diversion programs versus stricter sentencing laws as interventions for adolescent conduct problems, or mainstreaming behavior-disordered children versus alternative schools and pull-out programs. The sheer number of alternative programs necessitates the availability of sound evaluative data to determine which prevention programs should be used. This point is even more important in a climate of public skepticism about the cost-effectiveness of prevention in general and a time of federal cutbacks in budgets for social programs.

Past failures of prevention programs implemented on a large scale (such as some aspects of the community mental health movement), and even rare deleterious intervention effects (McCord, 1978, 1992), point to the need to determine that programs are effective before they are implemented on a large scale. Prevention programs are often very expensive, and ineffective programs can hinder the broader cause of prevention by deflating public confidence in the concept. Because of the long time lag between program implementation and observation of outcomes, program advocates may be unaware of deleterious effects and may unknowingly disseminate ineffective, or even iatrogenic, interventions.

It seems clear that the role of evaluation is critically important in the development and assessment of preventive interventions. As evaluation data play an increasingly central role in theory development, and as the number of programs competing for adoption and

use skyrockets, the need for high-quality evaluation data is more pressing than ever.

## Theory-Based Evaluation

Until about 15 years ago, evaluation science was concerned primarily with improving the validity of evaluation research via methodological improvements (Chen & Rossi, 1987). This focus was driven by the view that the development and use of adequate research methods was a necessary foundation for the science. Evaluators focused on design issues rather than theory development, and they were not necessarily even aware of how and why programs were expected to work (Wortman, 1983). The primary goal of evaluation was to conduct a methodologically rigorous study that provided internally valid information about whether the evaluated programs led to particular outcomes. This traditional orientation, while advancing evaluation science through methodological improvements, tended to produce atheoretical "black box" evaluations. Such studies are referred to as black box evaluations because they were conducted to determine the success or failure of the programs in their entirety (Bickman, 1987a; Chen & Rossi, 1983).

The reduction of programs to "black boxes" simplifies the use of experimental and quasi-experimental design methods (Lipsey & Pollard, 1989), but the black box design is not elaborate enough to answer questions about the specific effects of the evaluated programs (Cordray, 1986). For example, it is sometimes of interest to know whether some program components are more effective than others, or whether program effects increase with exposure ("dosage"). These issues become even more important when the black box method yields weak, null, or mixed results, because the evaluator cannot determine whether the poor results are due to theory failure or because the program has a mixture of effective and ineffective components (Chen & Rossi, 1980). As the development of effective research methods matured, theoretical concerns assumed a more central focus in evaluation science. Theory-based evaluation (Bickman, 1987b; Chen, 1989, 1991; Chen & Rossi, 1980, 1983, 1987, 1989) addresses the problems inherent in black box evaluation by linking the evaluation design to the program





theory. Bickman (1987a) defines program theory as "the construction of a plausible and sensible model of how a program is supposed to work." Understanding program theory requires opening up the black box and completely understanding the various functions of the program. Theory-based evaluation is designed to test multiple aspects of the program theory and to provide information about program specifics. This evaluation process is necessarily more complex than the traditional black box design and requires a thorough understanding of the program processes.

Theory-based evaluation is a relatively new process. As recently as a decade ago, fewer than 10% of the evaluation studies reviewed by Lipsey, Crosse, Dunkle, Pollard, and Stobart (1985) fully integrated program theory into the evaluation design. In addition to enhancing the quality and scope of evaluation efforts, articulation of a theory of prevention of a particular outcome forces program designers to think more clearly about the intervention. Crucial features of intervention can get prioritized. Expectations about when intervention effects should occur can heighten awareness among interventionists so that midstream modifications in intervention can occur.

Chen (1991) described six major domains of program theory. These domains provide a framework for understanding program functioning and facilitate the application of program theory to evaluation design. Analysis of each domain in the context of a specific program yields a component of the program theory for each domain. Each domain theory can then be used to formulate theory-based evaluation components corresponding to each domain. The evaluation is designed by selecting those evaluation components that will provide the desired information.

The *treatment domain* concerns understanding and documenting the conceptualization and design of the program treatment. Questions about the meaning, functioning, planning, and organization of the treatment are part of this domain. Social programs are often complex and difficult to deliver, so treatment that is planned may not be completely successfully implemented. For this reason, it is important that evaluations include strategies for documenting the treatment as actually implemented in addition to the treatment as planned. *Treatment theory* describes what the specific treatment should be; that is, what should be done to achieve the desired proximal results. *Treatment evaluation* describes the actual treat-

BEST COPY AVAILABLE

ment as planned, measures the treatment as implemented, and examines their congruence.

The *implementation environment domain* describes the environment in which the treatment should be implemented to be successful. Questions in this domain include logistic aspects of the program, such as whether the targeted group is actually exposed to the program, staff training, and delivery atmosphere (e.g., group sessions or individual training sessions). *Implementation environment theory* describes the atmosphere in which the specific treatment should be delivered to be effective. Accordingly, *implementation environment evaluation* describes the planned implementation environment, measures the actual implementation environment, and examines their congruence.

The *outcome domain* concerns the changes that occur as a function of the treatment program, both intended and unintended. Intended outcomes are those that the treatment was specifically designed to achieve, and they are typically measured and used to gauge program effectiveness. Unintended outcomes are effects of the treatment program that were not anticipated or predicted by treatment theory. They may be beneficial, benign, or detrimental. Theory-driven evaluations place a higher priority on measuring unintended outcomes than has been typical in traditional evaluation (Sherrill, 1984). A comprehensive *outcome theory* will explore both the types of outcomes suggested by the specific program treatment theory and any likely unintended outcomes. *Outcome evaluation* clearly identifies both types of outcomes in the specific program context.

The *impact domain* concerns assessing the impact of the treatment on the outcome. This domain involves a complete understanding of the relation between the treatment program and the outcome. Issues of causal inference and internal validity are also subsumed by this domain. Impact assessment is a primary focus of both traditional and theory-based evaluation designs, although theory-based evaluations may involve multiple treatment-outcome assessments. The hypothesized causal process in a specific program is specified in the *impact theory* of program functioning. The theory-based *impact evaluation* assesses the effect of the treatment on the outcomes, and differs from traditional impact assessments by employing a large evidence base for demonstrating program efficacy. Causal inference is strengthened by using randomized





experiments or multiple-method quasi-experimental designs (e.g., nonequivalent control group combined with interrupted time series pretests) where experiments are not possible. More evidence for program efficacy is obtained through the use of broader outcome definitions (including unintended outcomes) and the specification of intervening causal mechanisms.

The underlying causal processes that lead to treatment effects are the subject of consideration in the *intervening mechanism domain*. Program treatments are rarely expected to affect outcomes directly—more commonly the treatment effects operate through some intervening mechanism. *Intervening mechanism theory* specifies how these intervening mechanisms operate in the specific program, and *intervening mechanism evaluation* empirically confirms the hypothesized underlying causal process.

Finally, the *generalization domain* concerns how the results of the evaluation might be used in populations and settings in the future. *Generalization theory* explores how the specific program might operate in different populations or settings (e.g., in a different implementation environment). Assessing the function of the treatment program in different populations and environments is one way to perform a *generalization evaluation* of the specific program.

Theory-based evaluations have an emphasis on measuring program implementation and clearly specifying the causal mechanisms responsible for treatment impact. More information about causal explanation can be gathered in theory-based evaluations because they carefully identify and measure implementation, intervening mechanisms, and broader outcomes (Cook et al., 1993). Because of this emphasis, theory-based evaluations have the potential to provide additional information about the causal mechanisms operating in the illness process that is the target of the intervention.

The amount of information gathered in theory-based evaluations is much greater than in traditional evaluations, but the information comes at the cost of greater methodological complexity. The additional complexity of the theory-based evaluation process increases the need for varied research designs and analytical strategies (Lipsey & Pollard, 1989). A benefit of this complexity is that it has focused more attention on elaborating important components of the process, such as implementation (Palumbo & Oliverio, 1989) and strength of treatment (Scott & Sechrest, 1989).

It is important to note that although there are numerous potential advantages to using theory-based evaluation, it may not be possible in all evaluation environments. Bickman (1989) explains that increased costs and inadequate knowledge may sometimes preclude the use of theory-based evaluation. Compared to traditional evaluations, theory-based evaluations require more planning, a greater amount of measurement development, more extensive data collection, more complex data storage and analysis, and additional report writing activities. All are expensive, and the resulting increase in costs may exceed the available resources. Lack of technical knowledge (how to develop a program theory) and content knowledge (about the evaluated program) may also be a barrier to the use of theory-based evaluation. Lipsey and Pollard (1989) also note that the complex research designs and statistical analyses required for some theory-based evaluations may daunt some evaluators. Cordray (1989) discusses some limitations of theory-based evaluations in the policy context and proposes some alternative ways of achieving similar evaluation goals (e.g., multiple methodologies, meta-analysis).

## Developmental Theory and Evaluation in the Prevention Context

The prevailing theory of prevention holds that the inhibition of physical or mental illness is accomplished by reducing risk factors and increasing protective factors that are associated with disease or mental disorder (Robins, 1992). Risk factors increase the likelihood, and protective factors decrease the likelihood, that a disease or disorder will occur in a given individual relative to the general population. Accordingly, prevention research is largely concerned with the identification and elaboration of risk and protective factors, and preventive intervention efforts seek to decrease risk factors and increase protective factors that will inhibit the development of disease or mental disorder in individuals.

This view of prevention theory leads to two major theoretical and methodological implications for prevention research. First, theories of prevention must include causal chains where background risk and protective factors affect subsequent individual and environmental factors that ultimately lead to the occurrence or





prevention of a disorder. The effects of the background variables operate through the intervening variables to affect disorder outcomes. Prevention evaluation methods must allow for the formulation and testing of such mediated models of causality. Second, this ordering of events in the prevention process implies a temporally serial process where early risk and protective factors affect later mediators and even later disorder outcomes. Prevention evaluation must employ longitudinal methods where prevention processes can be followed through time to determine their impact on the ultimate development of the disorder. The demands of prevention evaluation are well served by a theory-based evaluation approach.

The need for increased theoretical and methodological precision in prevention research is acknowledged in the 1994 IOM report on preventive intervention. The report lists several obstacles to the progress of the field of mental health prevention, one of which is the lack of an organizing theoretical framework. Understanding the operation of risk and protective factors on proximal and distal mental disorder outcomes is especially underdeveloped. These theoretical gaps can be filled with information generated by well-designed evaluations of prevention programs.

The mediational and longitudinal aspects of prevention theory and research maximize the potential utility of theory-based evaluation of mental illness prevention programs, because theory-based evaluation focuses on intervening variables and multiple outcomes. Additional evidence for the utility of theory-based evaluation in the prevention context can be found in the 1994 IOM report, which lists six types of information needed to assess the quality of published research on preventive intervention programs: (a) a description of the hypothesized risk and protective factors; (b) a description of the targeted group; (c) a description of the intervention program itself; (d) a description of the research design and statistical methods used in the prevention study; (e) a description of the implementation of the preventive intervention; and (f) a description of whether outcomes have been influenced. The authors of the IOM report recommend that researchers, practitioners, and policymakers use these assessment criteria to make critical decisions about the mental illness prevention field. Theory-based evaluations of mental illness prevention programs can provide all six types of information.

Developmental theory has an important role in the establishment of mental illness prevention programs (Cicchetti & Toth, 1992; Sroufe & Rutter, 1984). Theory-based evaluations of prevention programs must also incorporate developmental theory, because designing a theory-based evaluation requires a complete understanding of program functioning. Acquiring such understanding in the prevention context requires taking a developmental theoretical perspective, because prevention theory assumes that early risk and protective factors affect later mental illness outcomes. In addition to providing information about the causal functioning of the prevention effort, theories about the developmental course of mental illness provide information critical to the design of preventive interventions. What characteristics identify the target group? At what age is intervention most effective? What potential preventive factors should be targeted, and when? What are potential short-term and long-term outcomes that can be used to determine the progress of the prevention effort? All of these questions can be informed by well-articulated developmental theories of mental illness etiology. Prevention evaluation research and developmental theory are mutually informative because both are primarily mediational and longitudinal in nature.

For example, Dodge (1986, 1993b) has outlined a model of the development of conduct disorder that includes mediational components. Dodge's model includes several intervening cognitive stages between background risk factors and the ultimate development of conduct disorder. Such intervening variables are important in the development of prevention evaluation because they can serve as targets of intervention and proximal outcomes that can demonstrate the midterm effectiveness of long-term prevention efforts.

Another example is the work of Sroufe and his colleagues (Sroufe, 1989, 1991), who have explored early pathways to psychopathology (such as attention deficit hyperactivity disorder) by comparing normal and pathological development. Such an approach leads to an understanding of where and when during development initial psychological deviations begin, and what characteristics denote the branching of a pathological developmental process from a normal one. This kind of theorizing can help prevention efforts by identifying the temporal sequence of events in the development of the disorder, and potential indicators of incipient mental illness. This information can inform the design of preventive interventions by





guiding the early identification of target groups and the timing of the interventions.

There is a wealth of basic research on conduct disorder (Yoshikawa, 1994), including several recent and ongoing longitudinal studies (Lahey et al., 1995; Zoccolillo, Pickles, Quinton, & Rutter, 1992). Enough basic research has been conducted and theoretical understanding has progressed to the point where it is reasonable to design and evaluate conduct disorder prevention programs (CPPRG, 1992; Guerra et al., 1995; Kellam, Rebok, Ialongo, & Mayer, 1994). Consistent with the current state of theoretical understanding of conduct disorder, these preventive interventions include longitudinal designs and focus on intervening developmental mechanisms.

Longitudinal preventive interventions, in addition to their potential for reducing mental illness, can be especially theoretically informative (Dodge, 1993a; Kellam & Rebok, 1992; Robins, 1992). Targeting hypothesized precursors of later illness for intervention can validate the underlying theory if influencing those precursors interrupts or precludes the development of the illness. Because of this, Robins (1992, p. 4) points out that prevention experiments are superior even to treatment experiments in terms of their value in providing information about the causal relation between preventive factors and outcomes. Kellam and Rebok (1992) agree with Robins that preventive trials enhance causal inference, and they identify a further benefit of prevention experiments. Citing the Woodlawn epidemiological studies (Kellam, Brown, Rubin, & Ensminger, 1983), Kellam and Rebok note that the development of mental illnesses is characterized by multiple background variables, multiple intervening variables, and multiple outcomes. There are typically several plausible developmental paths among these groups of variables—single background variables can affect multiple intervening variables, single intervening variables can affect multiple outcome variables, and so on. Prevention evaluations can help clarify the complicated relations among the variables in the developmental process. Specific intervening variables can be targeted by the preventive intervention, and the effects on outcomes evaluated. If the intervention is successful, observing which outcomes are affected by the intervention provides information about which outcomes are caused by the targeted intervening variables.

All of the conduct disorder prevention programs cited above have at least partially employed this developmental theory/prevention evaluation approach. Kellam and his colleagues have used this approach extensively to explore the effects of school achievement and aggressive behavior on the subsequent development of depression and delinquency in the Baltimore preventive trial. They have published many of these results, and the progress of their work to date serves as an excellent example of the developmental theory/prevention evaluation research process.

Kellam and Rebok (1992) note that the first step in using prevention evaluation results to inform developmental theory is to describe the developmental process the intervention is designed to affect. This understanding of the natural developmental process, what Kellam and Rebok call a *baseline developmental model* of mental illness, is required to get an adequate understanding of the underlying illness process. For example, Kellam et al. (1991) analyzed longitudinal data concerning the natural development of concentration problems, aggressive behaviors, and school achievement for a sample of children who did not receive any treatment or intervention. Using statistical methods sensitive to change over time, they found that in the absence of preventive efforts, concentration problems led to subsequent aggressive behavior and lower achievement. Earlier research (Kellam et al., 1983) had shown that aggressive behavior is a precursor of delinquency and that low achievement results in later depressive symptoms. What emerges from this research program is a picture of how the concentration problem's background variable influences the aggression and achievement intervening variables, that in turn affect the long-term outcomes of depression and delinquency. This baseline developmental model describes potential proximal targets for intervention (conduct problems, aggressive behavior, school achievement) and possible divergent pathways to different outcomes (aggression to delinquency and achievement to depression).

Research of this type has come to be called *developmental epidemiology* (Kellam & Rebok, 1992; Scott, Shaw, & Urbano, 1994), because of its focus on the development of naturally occurring mental illness. Developmental epidemiology is characterized by longitudinal research on untreated populations. Kellam et al.'s (1991) research fits this definition because it employed longitudinal methods to chart the development of delinquency and depression,





and because it focused on naturally occurring mental illness processes in untreated children.

The next step in the developmental theory/prevention evaluation process is to conduct an intervention that targets hypothesized risk or protective factors in the developmental model, and confirm that the intervention is successful. Building on their developmental epidemiological results, Kellam and his colleagues conducted and evaluated two preventive interventions designed either to decrease aggressive behavior or increase school achievement. The interventions were delivered to children in the first grade of public school, and each child received one of the interventions but not both. Following from their developmental epidemiological findings, they expected that decreasing the aggressive behavior risk factor or increasing the reading achievement protective factor would indirectly decrease the likelihood that children who participated in the prevention program in first grade would experience subsequent delinquency or depression. Aggressive behavior and reading achievement are the intervening variables in the developmental model. Results indicated that the prevention efforts were successful in affecting the proximal targets in the desired ways—the interventions reduced aggressive behavior and improved reading achievement in first-grade children (Dolan et al., 1993).

The final step in the developmental theory/preventive evaluation process is to confirm that successfully intervening on risk and protective factors has a preventive effect on mental illness outcomes. Kellam and his colleagues have found some preliminary evidence that the impact of the intervention on first-grade aggressive behavior and reading achievement does indeed prevent the later development of delinquency and depression, the outcomes according to the developmental model. Participation in the intervention designed to promote reading achievement was related to reductions in depressive stability and symptoms (Kellam, Rebok, Mayer, Ialongo, & Kalodner, 1994). Participating in the intervention designed to reduce aggressive behaviors in first grade was related to a reduction of aggressive behaviors at the time of transition to middle school, when behavior problems are typically more likely (Kellam, Rebok, Ialongo, & Mayer, 1994). These results suggest that the interventions do indeed have a preventive effect on later outcomes, but further follow-up is yet necessary to prove this definitively.



BEST COPY AVAILABLE

Taken together, the results generated by this program of research support the efficacy of the intervention and also provide information about the putative developmental model of delinquency and depression. The preventive interventions successfully reduced aggressive behavior and increased reading achievement, which in turn reduced subsequent aggressive behavior and depression. The combination of mediating variables and longitudinal data in this example provides strong support for the hypothesized developmental model of delinquency and depression, especially because the intervening risk and protective factors are shown to operate over time.

## Methodological Implications of the Developmental Theory/Prevention Evaluation Approach

Basing an empirical evaluation of a prevention trial on a theoretical model of development and an a priori concept of how intervention should lead to specific outcomes opens numerous methodological issues. Several pertinent issues will be discussed next.

*Randomization.* In the interest of testing both the effectiveness of an intervention and the veracity of a developmental model, random assignment of participants to prevention and control groups is crucial. Only through randomization can we be confident that the emergent differences between these two groups are due to the intervention and not prior selection factors. In practice, this simple goal turns out to be very difficult to achieve.

The appropriate method for assigning participants to conditions involves first identifying all participants and guaranteeing (as much as possible) their long-term participation. Then, random assignment should be made to prevention versus control conditions (or whatever additional groups are planned, such as partial prevention). However, few participants will react well to expecting to become part of an intervention and then being denied the opportunity. Likewise, some of the prevention group participants might not agree to receive intervention. If the prevention group includes only those persons who are willing to be involved in intervention, but the control group includes all participants, the design becomes





biased. It might be that "willing" participants are more likely to improve over time than "nonwilling" participants, even without intervention. Design issues also suggest that it might be biasing to inform participants in the control condition that they are "control subjects" because it could induce a negative expectation for later development. Ethical considerations lead some designers to reject this proposition altogether.

Some designers resolve this dilemma by identifying one group of individuals who would be willing to participate in an intervention and then identifying another group from a separate source who would be willing to participate in a nonintervention longitudinal study (without mentioning the intervention). This design resolves some ethical dilemmas, but it does not constitute random assignment because the two groups come from separate sources and are likely to differ on many dimensions a priori. An improvement would be to identify potential participants from one source, randomly assign them to receive prevention or not, and then approach them about participation. This approach retains the attribute of randomization, but it may lead to differential (or worse, biased) nonparticipation rates and attrition over time. Still a further improvement has been made in the FAST Track program run by the CPPRG (1992). In this program, participants were first approached to participate in a longitudinal study of child development. After consent was confirmed, participants were randomly assigned to prevention versus control conditions. Then, the prevention group participants were approached to participate in the prevention program. This design retains the virtues of random assignment as long as all of the prevention group participants, even those who refuse to participate in the intervention, are considered part of the intervention group sample for later evaluation. Thus, the prevention group condition comes to include both recruitment and intervention as part of its defining features.

*Clustering.* Another major problem is that interventions are often implemented in the context of an existing classroom, school, or community. The interventions often involve the teacher, the peer group, or the school itself. Random assignment of children within a classroom (or school) to intervention versus control conditions could lead to contamination of the control group children, because

they might indirectly benefit from intervention with the teacher or peers. Therefore, random assignment sometimes must occur with the classroom or school as the unit of assignment. The unit of assignment should correspond with the unit of analysis. A major problem that arises is that when the unit of analysis is the classroom or school, the number of cases for analysis is often small. A prevention program might involve hundreds of children but only 10 classrooms (or two schools), so the statistical power to detect differences between the intervention and control groups is often too weak, given the expected magnitude of effects of intervention. Evaluators sometimes try to avoid this problem by assuming that assignment to classroom or schools has been random and that interaction among children is nonexistent, so that the child can be used as the unit of analysis. Unfortunately, these assumptions are rarely met.

*Sampling Bias.* Even when initial random assignment has been achieved successfully, biases can creep into the samples over time, through differential participation, attrition, and mobility. Attrition is always a problem for longitudinal inquiry, and it is incumbent on the evaluator to test whether those children who leave differ in important substantive ways from those children who remain in the study. If children who leave differ, then the evaluation findings can be generalized only to the populations with characteristics of the remaining participants.

Differential attrition between the intervention and control groups is a second problem. Because of the attraction of receiving help and the development of commitment to interventionists, it might be that families in the intervention group are more likely to remain in the study. The opposite might also occur, if the intervention proves taxing or difficult to complete. In either case, if attrition rates differ for these two groups and if children who leave differ from remaining participants, then differences develop over time between the two groups even when the original groups had been randomly assigned and had not differed. These differences might take the form of systematic selection biases that are confounded with experimental conditions. The result is that the groups could appear to become different as a function of intervention, when the real difference can be attributed to an emergent selection bias.





There are no fully satisfactory ways of counteracting this problem, and evaluators must note this limit in their findings.

*Examining Subgroups.* Often, a preventive intervention will have different impacts on subgroups of children. For example, a program to prevent conduct problems might be effective with children who have minor behavioral difficulties but might not have any impact at all on children with initially severe behavior problems. Or an intervention might be effective for one gender but not the other, for one ethnic group but not another, or for one age group but not another. Analyses of intervention effects on subgroups are potentially very exciting for the field because they potentiate the matching of different kinds of interventions with different kinds of children.

One common mistake made by evaluators is to determine subgroups within the intervention group and then contrast each subgroup with the entire control group. Thus, the "mild" behavior-problem subgroup within the intervention group might be contrasted with the entire control group (including both "mild" and "severe" behavior-problem control subgroups). This obvious mistake can be avoided by assigning children to subgroups based on preintervention characteristics only, and then contrasting intervention versus control condition within subgroups only.

Another popular analytic technique is to examine subgroups of intervention children who meet different characteristics across the course of intervention. For example, one might examine the effects of intervention only on that subgroup of intervention children who complete every session of intervention (or who show interest in the program). This subgroup is often contrasted with the entire control group of children, a problem that has been noted above. One problem here is that there is no satisfactory way of determining within the control group which children also would have completed every session (or shown interest), so that there is no way to compose an appropriate comparison group. Some attempts might be made to simulate this subgroup by empirically predicting participation (or interest) in the intervention based on preintervention characteristics, and then using this regression-type prediction to form a similar control subgroup based on preintervention characteristics. This approach has merit, but it must be noted that it is never fully



BEST COPY AVAILABLE

adequate because the empirical prediction has been conducted after the fact and might not replicate. Usually, the evaluator's only option is to report findings for subgroups as a nonexperimental and post hoc finding.

## How to Evaluate a Long-Term Intervention Before the Long Term Is Over

Many programs aim to prevent long-term maladaptive outcomes, such as adolescent delinquency, drug use, or teen pregnancy, that will not be evident for many years. Other persons and agencies often do not have the luxury to wait many years before deciding whether to implement a promising program. When a prevention program is based on a theoretical developmental model, with both proximal and distal goals, it may be possible to acquire initial evaluation evidence even in the short term.

Intervention program effects can be divided into three categories of implementation, proximal effects, and distal effects. The first level is implementation. Evaluators can determine whether it is possible to implement an intervention as planned by evaluating whether the proposed intervention actually takes place. For example, a program designed to deliver parenting training to high-risk mothers can be evaluated in terms of the percentage of mothers who participate, who attend sessions, and who receive intervention services.

The second level is proximal effects. Most prevention programs attempt to alter some characteristic in the child or environment that, in turn, is hypothesized to lead to later changes in the outcome of interest. For example, the FAST Track program (CPPRG, 1992) is an attempt to alter young children's academic and social skills, based on the hypothesis that proximal changes in these skills will lead to distal changes in adolescent juvenile court adjudication and conduct disorder. Evaluators can determine whether the intervention indeed alters the proximal process as hypothesized. It is unlikely that a program that fails to affect proximal change when this change is hypothesized will show long-term, sleeper effects on the distal outcome. Thus, programs that propose possible implementation and proximal effects should be able to demonstrate those effects even in a short period of time.





Of course, positive implementation and positive proximal effects do not guarantee long-term effects on the distal measures of interest. Thus, the last level of evaluation is the distal, ecologically valid level. This measure is what is being prevented.

## Evaluation of Dissemination

The final stage of the prevention research cycle (IOM, 1994) is the wide-scale dissemination of a laboratory-based program into local communities. Instead of a program being designed, implemented, and evaluated by a group of scholars or innovators (as with most prevention research), the program comes to be owned by local communities (as with most interventions). This step is necessary, of course, for the new program to become part of the fabric of the community and to continue to be implemented across time. One important measure of a program's worth is the degree to which it becomes owned by local communities.

An emerging field within prevention research (Weissberg, 1990) concerns the generation and evaluation of theories, strategies, and policies to apply broad-scale dissemination of programs (e.g., how effective are demonstration programs versus advertising in leading to wide-scale implementation of model programs?). A separate question concerns the evaluation of those disseminated programs (e.g., is a program that has been disseminated to a local community as effective in preventing negative outcomes as the model program that spawned the dissemination?). Each of these activities requires sound evaluation efforts.

*Evaluation of the Implementation of Disseminated Programs.* Weissberg (1990) has noted the debate among community psychologists regarding the best way to implement social innovations in a new context. The profidelity perspective is that the disseminated program should be adopted with all features of the original model being retained. Only by doing so can local officials be confident that the disseminated program will achieve the same outcomes as the model program. The proadaptation perspective is that because each local community has a different set of resources, culture, and needs, a model program must be changed and adapted to serve the local culture. The original program was

developed within a context, and the disseminated program must also be implemented within its local context. Weissberg (1990) has argued that optimal disseminations will integrate features of each perspective.

Evaluation of the dissemination process will require program developers to be clear about which features of a model program are crucial to a theory of prevention and which features have been included to meet the needs of a local culture. In the FAST Track program (CPPRG, 1992), for example, training parents in behavior management skills is a crucial part of the prevention theory, but administering this skill training in groups over an evening meal versus administering it in Saturday morning classes is less crucial to the prevention theory and more crucial to local norms and ways of conducting business. Perhaps the crucial feature of the latter question is the more general point that the intervention be administered in a setting that is comfortable for parents. Program developers must be clear about the conceptual features of implementation that are important.

Once program developers are clear about which features must be disseminated rigorously and faithfully, then evaluation of a dissemination effort will follow from this set of guidelines.

*Evaluation of the Effectiveness of Disseminated Programs.* Because of the differences between original and disseminated programs in participants, staff, and implementation (including both planned and unplanned differences), it is imperative that the effectiveness of disseminated programs also be evaluated. Unfortunately, many model programs lose their effectiveness when disseminated (Haskins, 1989). One of the methodological dilemmas for prevention researchers concerns the strategies that can be used to evaluate a locally owned program.

By definition, disseminated programs are not administered by prevention researchers, so the opportunity for rigorous research designs and data collection may be minimal. Still, local disseminators can conduct important evaluations of their own programs. In fact, we would argue that local disseminators should be required to evaluate their own efforts, in the same way that a local practitioner of health or mental health interventions should be required to evaluate her or his interventions. These evaluations are best conducted with randomized experiments, but, of course, such





designs alter the meaning of full dissemination. Alternatives include the use of single-subject case designs, in which a program is implemented for a period of time. Outcome measures, collected by local program owners, are contrasted for the period prior to implementation and afterward. If local disseminators lack confidence that differences are due to the dissemination (rather than secular changes in a measure), then the program could be temporarily stopped and reimplemented to chart changes as a function of the program implementation. At the very minimum, local program administrators should collect outcome measures for a program, to compare performance of a community with performance of its model or other communities.

## Statistical Issues in Prevention Evaluation

Cook and Shadish (1994) have noted that social experiments combining theoretical testing with field demonstrations of innovative programs, like the kind of prevention evaluations we are presenting here, are becoming increasingly more common. The increased prevalence of theory-based research has raised a concomitant need for statistical methods capable of appropriately analyzing the data produced in more complicated research designs. It is difficult to review statistical methods for prevention evaluations because the theory-based approach is relatively new, and many statistical techniques that are applicable to prevention evaluations are currently in development. At this point, any review is sure to be out of date by the time it is published. However, several established statistical procedures are useful in prevention evaluation, and a few developing trends are worth watching.

Developmental theory usually specifies a causal chain of variables in the developmental course. Interventions that arise out of developmental theory typically target variables in early stages of the causal chain that are presumed to inhibit the later development of undesired outcomes. Prevention evaluation designs require complex statistical methods that can test the effects of intervening variables, and appropriately model longitudinal outcome processes. The analysis of data generated by preventive interventions is further complicated by the fact that the targeted outcomes often occur infrequently in the general population (i.e., they are *low base-rate*

BEST COPY AVAILABLE

phenomena). An underlying assumption of many statistical techniques is that the measured outcome variable is normally distributed. The distribution of a low base-rate phenomenon in the population is not normal, so an associated sample outcome measure will not be normally distributed. Even when an outcome is normally distributed in the population, sample outcome measures collected in the field tend not to be normally distributed (Micceri, 1989). It is critical for the prevention researcher to be aware of statistical tools that are appropriate for (a) detecting the effects of intervening variables, (b) modeling longitudinal processes, and (c) nonnormally distributed outcome variables.

*Intervening Variables.* Intervening variables can function as mediators or moderators of the effect of the background variable on the outcome. Mediation occurs when the effect of the background variable on the outcome operates through the intervening variable. Change in the background variable is associated with change in the intervening variable, which in turn is associated with change in the outcome variable. In a fully mediated relationship, there is no direct effect of the background variable on the outcome variable after the intervening variable is accounted for; all of the effect operates through the intervening variable. Moderation occurs when the intervening variable changes the relationship between the background variable and the outcome variable. An intervention effect may operate as a moderating variable by changing the association between a background variable and an undesired outcome.

It is important to note that the relationships involving an intervening variable must be carefully specified on the basis of an appropriate theoretical model before they are statistically tested. It is difficult, if not impossible, to distinguish between mediated and moderated relationships on the basis of statistical results alone. It is especially important to have a clear understanding about any potential moderator effects in the intervention process. If undetected, moderated relationships can obscure the effects of the intervention on the outcome (Bryk & Raudenbush, 1988).

It may be necessary to test both mediated and moderated relations in the context of the same prevention evaluation. For example, recall that Kellam and his colleagues first established a mediated relationship showing that increased concentration problems (the background variable) were associated with increased aggressive behavior (the intervening variable), which was in turn associated with decreased academic achievement. Next, they tested an intervention where the good behavior game was intended to break the causal chain between concentration problems and decreased academic achievement. In this case, the intervention was an intervening variable that moderated the relationship between concentration problems and decreased academic achievement. Increased concentration problems were associated with decreased academic achievement for those who did not receive the intervention, but there was no such relationship for those who did receive the intervention—the intervention changed the relationship between the background variable and the outcome variable.





Hoyle and Smith (1994) point to the utility of structural equation modeling (SEM) in modeling mediated relationships among variables. Bollen's text *Structural Equations With Latent Variables* (1989) is a good reference for understanding SEM techniques. In cases where SEM is not feasible, Baron and Kenny (1986) outline a regression procedure for detecting mediated relationships.

Moderated relationships are tested by detecting the strength of statistical interactions. There are several approaches for investigating statistical interactions, both in the context of the analysis of variance (Judd, McClelland, & Culhane, 1995; Rosnow & Rosenthal, 1989) and regression (Aiken & West, 1991; Jaccard, Turrisi, & Wan, 1990). Be advised that detecting statistical interactions is difficult, and there is disagreement about how best to do it (Meyer, 1991).

Multisample analysis in SEM (Bollen, 1989) is also useful for detecting moderated relations. In multisample analysis, relationships in one subgroup of a defined population (e.g., those who received intervention) are contrasted against the same relationships in another subgroup (those who did not receive intervention).

*Longitudinal Data.* Preventive interventions are aimed at an early point in the development of an undesired outcome and designed to inhibit the subsequent development of that outcome. Outcomes are measured on the same individuals across time so that differences between the developmental trajectories of the intervention and control groups can be detected. Measurement of the same individuals over time causes errors from a particular measurement

occasion to be correlated with the errors from preceding occasions of measurement. The statistical tools used to analyze such data must be sensitive to changes in developmental trajectory, and they must be able to handle data from correlated measurement occasions without generating biased results.

Hierarchical linear modeling (HLM), when used to estimate growth curve models, meets both of these conditions. Growth curve modeling captures information about developmental trajectories because the model estimates a growth curve for each subject. The parameters describing growth curves for individuals in different groups can be tested for differences. The rates of change in the outcome variable among groups are directly compared, allowing a very sensitive test of change over time. Growth curve models also incorporate the correlated errors of multiple-occasion outcome measurements, so the results of growth curve analysis are not biased by correlated measurement error.

The potential of growth curve models for analyzing data from prevention evaluations, where the preventive intervention is designed to alter the developmental trajectory of an illness process, is enormous. The text *Hierarchical Linear Models* (Bryk & Raudenbush, 1991) is an excellent introduction to the general hierarchical linear model, and it illustrates how HLM can be used to estimate the growth curve model. Some growth curve models can also be estimated in SEM (Farrell, 1994; Muthén, 1991; Willett & Sayer, 1993).

Original conceptualizations of HLM were limited to linear models of growth trends. While linear growth curve models including power terms reasonably approximate several common nonlinear growth trends (Burchinal & Appelbaum, 1991), the development of statistical technologies to estimate growth curve models with nonlinear growth trends is advancing rapidly (Hedeker & Gibbons, 1994; Liang & Zeger, 1986; Lindstrom & Bates, 1990; Zeger & Karim, 1991; Zeger & Liang, 1986).

*Nonnormally Distributed Outcome Variables.* As noted above, outcome variables in prevention evaluations are sometimes nonnormally distributed, and therefore violate the assumptions of many statistical procedures. One way to treat nonnormally distributed variables is to transform them (e.g., by taking their log or square root). Such transformations can normalize the distribution of non-





normally distributed raw variables, making them more appropriate for analysis using statistical methods with normal-distribution assumptions (Judd et al., 1995). Data transformations may not always be possible or advisable, however. The distribution of the transformed variable may not be sufficiently normalized as to be suitable for analysis by normal-distribution methods. Another drawback to transformations is that the scale of a raw variable is changed by transformation, often requiring conversion back to the original metric before the results of the analysis can be interpreted. Such a conversion may not be possible in multivariate models, or in models where both outcome and predictor variables require transformation.

In some cases, it may be more useful to use a statistical technique that is specifically designed to handle nonnormal data, such as generalized linear modeling (McCullagh & Nelder, 1989). Generalized linear models have underlying assumptions just like normal-distribution techniques, and they are not a panacea for analyzing nonnormal outcome variables. The analyst must have a good understanding of the relationship between the mean and variance of the outcome variable to specify an appropriate generalized linear model.

Many well-developed statistical techniques exist for cases where the outcome variables are categorical, such as logistic regression for binary outcomes and Poisson regression of counts (Agresti, 1990; McCullagh & Nelder, 1989; Muthén, 1984). Such models are useful for analyzing data from developmental epidemiological studies (Scott et al., 1994), where the outcome variables are counts of at-risk individuals who do or do not develop the undesired outcome. Models for categorical variables may also be useful for cases where the distribution of a continuous outcome variable is exceedingly difficult to accommodate through transformation or generalization. In general, information is lost when a continuous outcome variable is categorized, but a radical solution for very ill-conditioned continuous data might be to divide the continuous variable into categories, then analyze the resulting categorical variable.

One promising approach to the problem of ill-conditioned data focuses on the measurement of outcomes rather than the statistical modeling of those outcomes. Sum or mean scores composed of groups of items with poor distributions are usually equally poorly

BEST COPY AVAILABLE

distributed. An alternative method of deriving scores is proposed by Thissen (1993), who explains how item-response models can be used to generate comprehensive, stable scores from sets of poorly distributed individual items, or even groups of items on markedly different scales.

## Conclusion

The nation's health agenda targets the reduction of several unhealthy behaviors that have proven to be resistant to intervention with adults: alcohol, tobacco, and other drug use; irresponsible sexual behavior; and violent and abusive behavior. While these behaviors are difficult to improve via intervention with adults, much promise exists in the prospect of targeting interventions designed to prevent the development of these behaviors in children. Strategies of prevention require an understanding of not only the static features of unhealthy behaviors but also the way they develop over time. Prevention evaluation methods must test the ability of preventive interventions to interrupt the development of undesired behaviors early, and how well they inhibit the subsequent development of unhealthy behaviors over time. We have described a research approach that is capable of serving both needs—testing prevention program efficacy over the entire course of development, and providing critical information about the specifics and predictions of etiological theories.

Evaluation research can be a useful tool fostering the development and implementation of effective intervention programs. We feel that the approach to the evaluation of prevention programs that we have presented in this chapter can help achieve the nation's public health goals at several stages in the prevention research cycle:

1. *Theory development.* The interplay of developmental theory and evaluation research methods will allow critical tests of specific theoretical predictions in naturalistic settings, helping to develop and refine our theoretical understanding of illness processes.
2. *Implementation clarity.* Rigorous evaluation research demands require that the program implementation strategy be fully articulated and that the actual program delivery be well measured. Such thoroughness will promote the feasibility of experimental interventions and the accuracy of disseminated versions, ultimately enhancing the effectiveness of large-scale prevention programs.
3. *Identification of effective programs.* Experimental and quasi-experimental evaluation research designs will provide accurate information about the efficacy of intervention programs, allowing the identification of the most promising of numerous, potential programs. Promoting and implementing only the most effective intervention programs will ensure that public health prevention efforts will have maximum impact.
4. *More effective disseminated programs.* Good evaluative data will allow the practitioner to review a range of options before choosing an effective program that is feasible in his or her implementation environment. Providing objective information that allows practitioners to choose optimum programs will enhance the likelihood of success of those programs. The generalization evaluation component of the theory-based evaluation model specifically emphasizes the testing of intervention programs in various environments, placing emphasis on the evaluation of disseminated interventions.
5. *Grounds for improvement.* An important feature of the prevention research cycle is that it is iterative. High-quality evaluations will accurately document both the successes and the shortcomings of experimental intervention programs. This knowledge will guide the refinement of subsequent preventive interventions that will serve to improve the health of our nation's people.





## References

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173-1182.

Bickman, L. (1983). The evaluation of prevention programs. *Journal of Social Issues, 39,* 181-194.

Bickman, L. (1987a). The functions of program theory. In L. Bickman (Ed.), *Using program theory in evaluation: New directions for program evaluation* (pp. 5-18). San Francisco: Jossey-Bass.

Bickman, L. (Ed.). (1987b). *Using program theory in evaluation: New directions for program evaluation.* San Francisco: Jossey-Bass.

Bickman, L. (1989). Barriers to the use of program theory. *Evaluation and Program Planning, 12,* 387-390.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley.

Boruch, R. F., & Wothke, W. (Eds.). (1985). *Randomization and field experimentation: New directions for program evaluation* (No. 28). San Francisco: Jossey-Bass.

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin, 104,* 396-404.

Bryk, A. S., & Raudenbush, S. W. (1991). *Hierarchical linear models.* Newbury Park, CA: Sage.

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development, 62,* 23-43.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

Chen, H. T. (1989). The conceptual framework of the theory-driven perspective. *Evaluation and Program Planning, 12,* 391-396.

Chen, H. T. (1991). *Theory-driven evaluation.* Newbury Park: Sage.

Chen, H. T., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces, 59,* 106-122.

Chen, H. T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7,* 283-302.

Chen, H. T., & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10,* 95-103.

Chen, H. T., & Rossi, P. H. (1989). Issues in the theory-driven perspective. *Evaluation and Program Planning, 12,* 299-306.

Cicchetti, D., & Toth, S. L. (1992). The role of developmental theory in prevention and intervention. *Development and Psychopathology, 4,* 489-493.

Coie, J. D., & Dodge, K. A. (in press). Aggression and antisocial behavior. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development.* New York: John Wiley.

Conduct Problems Prevention Research Group. (1992). A developmental and clinical model for the prevention of conduct disorder: The FAST Track program. *Development and Psychopathology, 4,* 509-527.

Cook, T. D., Anson, A. R., & Walchli, S. (1993). From causal description to causal explanation: Improving three already good evaluations of adolescent health problems. In S. G. Millstein, A. C. Petersen, & E. O. Nightingale (Eds.), *Promoting the health of adolescents: New directions for the twenty-first century* (pp. 339-374). Oxford: Oxford University Press.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand-McNally.

Cook, T. D., & Shadish, W. R. (1994). Social experiments: Some developments over the past fifteen years. *Annual Review of Psychology, 45,* 545-580.

Cordray, D. S. (1986). Quasi-experimental analysis: A mixture of methods and judgment. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis: New directions for program evaluation* (No. 31, pp. 9-27). San Francisco: Jossey-Bass.





Cordray, D. S. (1989). Optimizing validity in program research: An elaboration of Chen and Rossi's theory-driven approach. *Evaluation and Program Planning, 12,* 379-385.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs.* San Francisco: Jossey-Bass.

Dodge, K. A. (1986). A social information processing model of social competence in children. In M. Perlmutter (Ed.), *Minnesota Symposia on Child Psychology* (Vol. 18, pp. 77-125). Hillsdale, NJ: Lawrence Erlbaum.

Dodge, K. A. (1993a). The future of research on the treatment of conduct disorder. *Development and Psychopathology, 5,* 311-319.

Dodge, K. A. (1993b). Social-cognitive mechanisms in the development of conduct disorder and depression. *Annual Review of Psychology, 44,* 559-584.

Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., Mayer, L. S., Laudolff, J., Turkkan, J. S., Ford, C., & Wheeler, L. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology, 14,* 317-345.

Farrell, A. D. (1994). Structural equation modeling with longitudinal data: Strategies for examining group differences and reciprocal relationships. *Journal of Consulting and Clinical Psychology, 62,* 477-487.

Fleming, M. (1996). *Healthy youth 2000: A mid-decade review.* Chicago: American Medical Association, Department of Adolescent Health.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches.* San Francisco: Jossey-Bass.

Guerra, N. G., Huesmann, L. R., Tolan, P. H., Van Acker, R., & Eron, L. D. (1995). Environmental stress and individual beliefs as correlates of economic disadvantage and aggression: Implications for preventive interventions among inner-city children. *Journal of Consulting and Clinical Psychology, 63,* 518-528.

Haskins, R. (1989). Beyond metaphor: The efficacy of early childhood education. *American Psychologist, 44,* 274-282.

Hedeker, D. R., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics, 50,* 933-944.

Hinshaw, S. P. (1994). Conduct disorder in childhood: Conceptualization, diagnosis, comorbidity, and risk status for antisocial functioning in adulthood. In D. C. Fowles, P. Sutker, & S. H. Goodman (Eds.), *Progress in experimental personality and psychopathology research. Special focus on psychopathology and antisocial personality: A developmental perspective* (pp. 3-41). New York: Springer.

House Select Committee on Children, Youth, and Families. (1988). *Opportunities for success: Effective programs for children, update 1988.* Washington, DC: U.S. Government Printing Office.

Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural equation models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62,* 429-440.

Institute of Medicine. (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research.* Washington, DC: National Academy Press.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

BEST COPY AVAILABLE

Judd, C. M., McClelland, G. H., & Culhane, S. E. (1995). Data analysis: Continuing issues in the everyday analysis of psychological data. *Annual Review of Psychology, 46*, 433-465.

Kellam, S. G., Brown, C. H., Rubin, B. R., & Ensminger, M. E. (1983). Paths leading to teenage psychiatric symptoms and substance abuse: Developmental epidemiological studies in Woodlawn. In S. B. Guze, F. J. Earls, & J. E. Barrett (Eds.), *Childhood psychopathology and development* (pp. 17-51). New York: Raven.

Kellam, S. G., & Rebok, G. W. (1992). Building developmental theory through epidemiologically based preventive intervention trials. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 162-194). New York: Guilford.

Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry, 35*, 259-281.

Kellam, S. G., Rebok, G. W., Mayer, L. S., Ialongo, N., & Kalodner, C. R. (1994). Depressive symptoms over first grade and their response to a developmental epidemiologically based preventive trial aimed at improving achievement. *Development and Psychopathology, 6*, 463-481.

Kellam, S. G., Werthamer-Larsson, L., Dolan, L., Brown, C. H., Mayer, L., Rebok, G. W., Anthony, J. C., Laudolff, J., Edelsohn, G., & Wheeler, L. (1991). Developmental epidemiologically based preventive trials: Baseline modeling of early target behaviors and depressive symptoms. *American Journal of Community Psychology, 19*, 563-584.

Lahey, B. B., Loeber, R., Frick, P. J., Hart, E. L., Applegate, B., Zhang, Q., Green, S. M., & Russo, M. F. (1995). Four-year longitudinal study of conduct disorder in boys: Patterns and predictors of persistence. *Journal of Abnormal Psychology, 104*(1), 83-93.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrica, 73*, 13-22.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometries, 46*, 673-687.

Lipsey, M. W., Crosse, S., Dunkle, J., Pollard, J., & Stobart, G. (1985). Evaluation: The state of the art and the sorry state of the science. In D. S. Cordray (Ed.), *Utilizing prior research in evaluation planning: New directions for program evaluation* (No. 27, pp. 7-28). San Francisco: Jossey-Bass.

Lipsey, M. W., & Pollard, J. A. (1989). Driving toward theory in program evaluation: More models to choose from. *Evaluation and Program Planning, 12*, 317-328.

Lorion, R. P. (1983). Evaluating preventive interventions: Guidelines for the serious social change agent. In R. D. Felner, L. A. Jason, J. N. Moritsugu, & S. S. Farber (Eds.), *Preventive psychology* (pp. 251-268). Oxford: Pergamon.

Lorion, R. P., & Lounsbury, J. (1981). Conceptual and methodological considerations in evaluating preventive interventions. In W. R. Tash & G. Stahler (Eds.), *Innovative approaches to mental health evaluation*. New York: Academic Press.

McCord, J. (1978). A thirty-year follow-up of treatment effects. *American Psychologist, 33*, 284-291.





McCord, J. (1992). The Cambridge-Somerville study: A pioneering longitudinal experimental study of delinquency prevention. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 196-206). New York: Guilford.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.

Meyer, D. L. (1991). Misinterpretation of interaction effects: A reply to Rosnow and Rosenthal. *Psychological Bulletin, 110*, 571-573.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115-132.

Muthén, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins & J. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 1-17). Washington, DC: American Psychological Association.

Palumbo, D. J., & Oliverio, A. (1989). Implementation theory and the theory-driven approach to validity. *Evaluation and Program Planning, 12*, 337-344.

Pearsol, J. A. (1987). Justifying conclusions in naturalistic evaluations. *Evaluation and Program Planning, 10*, 307-308.

Price, R. H., & Smith, S. S. (1985). *A guide to evaluating prevention programs in mental health* (DHHS Publication No. ADM 85-1365). Washington, DC: U.S. Government Printing Office.

Robins, L. N. (1992). The role of prevention experiments in discovering causes of children's antisocial behavior. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 3-18). New York: Guilford.

Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin, 105*, 143-146.

Scott, A. G., & Sechrest, L. (1989). Strength of theory and theory of strength. *Evaluation and Program Planning, 12*, 329-336.

Scott, K. G., Shaw, K. H., & Urbano, J. C. (1994). Developmental epidemiology. In S. I. Friedman & H. C. Haywood (Eds.), *Developmental follow-up* (pp. 351-374). San Diego, CA: Academic Press.

Sherrill, S. (1984). Identifying and measuring unintended outcomes. *Evaluation and Program Planning, 7*, 27-34.

Sroufe, L. A. (1989). Pathways to adaptation and maladaptation: Psychopathology as developmental deviation. In D. Cicchetti (Ed.), *Rochester Symposia on Developmental Psychopathology, 1*, 13-40.

Sroufe, L. A. (1991). Considering normal and abnormal together: The essence of developmental psychopathology. *Development and Psychopathology, 2*, 335-347.

Sroufe, L. A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development, 55*, 17-29.

Thissen, D. (1993). Repealing rules that no longer apply to psychological measurement. In N. Frederiksen, R. J. Misleavy, & I. Behar (Eds.), *Test theory for a new generation of tests* (pp. 79-97). Hillsdale, NJ: Lawrence Erlbaum.

U.S. Department of Health and Human Services, Public Health Service. (1991a). *Healthy children 2000*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services, Public Health Service. (1991b). *Healthy people 2000: National health promotion and disease prevention objectives* (DHHS Publication No. PHS 91-50212). Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services, Public Health Service. (1995). *Healthy people 2000: Midcourse review and 1995 revisions*. Washington, DC: U.S. Government Printing Office.

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts* (pp. 65-92). Washington, DC: Aspen.

Weissberg, R. P. (1990). Fidelity and adaptation: Combining the best of both perspectives. In P. Tolan, C. Keys, F. Chertok, & L. Jason (Eds.), *Researching community psychology: Issues of theory and methods* (pp. 186-189). Washington, DC: American Psychological Association.

Willett, J. B., & Sayer, A. G. (1993). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*.

Wortman, P. M. (1983). Evaluation research: A methodological perspective. *Annual Review of Psychology, 34*, 223-260.

Yoshikawa, H. (1994). Prevention as cumulative protection: Effects of early family support and education on chronic delinquency and its risks. *Psychological Bulletin, 115*, 28-54.

Zeger, S. L., & Karim, M. R. (1991). Generalized linear models with random effects; A Gibbs sampling approach. *Journal of the American Statistical Association, 86*(413), 79-86.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics, 42*, 121-130.

Zoccolillo, M., Pickles, A., Quinton, D., & Rutter, M. (1992). The outcome of childhood conduct disorder: Implications for defining antisocial personality disorder and conduct disorder. *Psychological Medicine, 22*, 971-986.

BEST COPY AVAILABLE